УДК 004.412:519.237.5

НЕЛІНІЙНІ РЕГРЕСІЙНІ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ ЯКОСТІ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА JAVA, ЩО СТВОРЮЮТЬСЯ З ВИКОРИСТАННЯМ ФРЕЙМВОРКУ HIBERNATE

Пухалевич А.В., к.т.н., доц.[1]; Коваленко В.В.[2]
[1,2] Національний університет кораблебудування імені адмірала Макарова
[1,2] Україна, Миколаїв
[1] andrii.pukhalevych@nuos.edu.ua; [2] vk.130703@gmail.com

Анотація. *В роботі, використовуючи перетворення Бокса-Кокса, побудовано нелінійні регресійні моделі RFC(CBO, WMC), CBO(WMC, RFC) та WMC(RFC, CBO). При побудові застосовувались емпіричні дані з 70 проєктів, доступних на GitHub.*
Ключові слова: *оцінювання якості розробки; нелінійна регресійна модель; фреймворк hibernate.*

Вступна частина. У сучасному світі Java продовжує утримувати провідні позиції серед мов програмування завдяки своїй інноваційності, адаптивності та продуктивності. Згідно з цьогорічним звітом компанії Azul, майже 70% із понад двох тисяч опитаних IT-фахівців у всьому світі зазначили, що більше половини їхніх застосунків створено за допомогою Java або працюють на Java Virtual Machine (JVM) [1]. Враховуючи широку популярність Java, а також розповсюджене використання фреймворку Hibernate – одного з найпоширеніших інструментів для роботи з ORM (Object-Relational Mapping), створення достовірних засобів оцінювання якості відповідного програмного забезпечення є актуальною задачею.

Оцінювання якості програмного забезпечення є складним завданням інженерії програмного забезпечення. Для вирішення цієї задачі можуть використовуватися нелінійні регресійні моделі, побудовані на основі метрик RFC, CBO та WMC [2–4]. Одним із підходів до оцінювання якості застосунків з відкритим кодом є метод [5], який передбачає обчислення меж довірчих та прогнозних інтервалів нелінійних регресій для показників RFC, CBO та WMC на системному рівні для тривимірних точок даних.

Оскільки створені раніше моделі будувались для оцінювання якості застосунків на інших мовах програмування або з використанням інших фреймворків, вони можуть призводити до зниження достовірності оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate.

Метою дослідження є побудова нелінійних регресійних моделей для оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate.

Основна частина. Для побудови трьох нелінійних регресійних моделей RFC(CBO, WMC), WMC(RFC, CBO) та CBO(WMC, RFC), якими оцінюватимуться відповідні метрики застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate, на GitHub було зібрано емпіричні дані з 70 проєктів аналогічної тематики. Вибірку було поділено на навчальну та тестову: 64% вибірки (45 застосунків) було відібрано для побудови моделей, а решту 36% (25 застосунків) – для перевірки їх якості.

Спочатку зібрані багатовимірні дані було перевірено на нормальність закону розподілу тестом Мардія, що базується на розрахунку параметрів багатовимірної асиметрії та ексцесу. У випадку, якщо хоча б одна з тестових статистик перевищує критичний рівень – розподіл

вважається не гаусівським, а дані мають бути нормалізовані для побудови регресійних моделей. Так як розраховане значення асиметрії перевищило критичне значення: 31,7609 > 25,1882, було прийнято рішення про необхідність нормалізації емпіричних даних [1].

Для нормалізації було використано одновимірне перетворення Бокса-Кокса для кожного параметру. Це дозволило побудувати лінійні регресійні моделі для нормалізованих значень метрик у кілька кроків.

Після виконання нормалізації першим кроком була перевірка відповідності перетворених даних нормальному закону розподілу. Якщо розподіл виявлявся не нормальним, із вибірки вилучався один рядок із найбільшим значенням квадрата відстані Махаланобіса, після чого перевірка повторювалась.

Далі, щоб виявити можливі викиди, було застосовано тест, що ґрунтується на квадраті відстані Махаланобіса для нормалізованих даних. У разі, коли для певного спостереження ця відстань перевищувала критичний поріг, з вибірки видалявся рядок із максимальним значенням квадрата цієї відстані, після чого побудова моделей розпочиналась з першого етапу.

Наступним кроком було визначення оцінок параметрів лінійної регресійної моделі та перевірка нормальності розподілу випадкової похибки за критерієм Пірсона.

Побудова лінійних регресійних моделей для нормалізованих даних відбувалася за формулою: $Z_Y = b_0 + b_1 Z_{X_1} + b_2 Z_{X_2} + \varepsilon$, де b_0, b_1, b_2 – оцінки параметрів лінійного рівняння регресії, отримані методом найменших квадратів, а ε – випадкова похибка.

Відповідно до описаних кроків, при побудові лінійних регресійних моделей було виявлено відсутність викидів у загальній вибірці емпіричних даних.

Для моделі RFC(CBO, WMC) було отримано наступні значення оцінок параметрів лінійного рівняння регресії: $b_0 = -0{,}3695$, $b_1 = 0{,}2539, b_2 = 1{,}2079$.

Для моделі CBO(WMC, RFC) було отримано наступні значення оцінок параметрів лінійного рівняння регресії: $b_0 = 4{,}3834$, $b_1 = -0{,}3661, b_2 = 0{,}2870$.

Для моделі WMC(RFC, CBO) було отримано наступні значення оцінок параметрів лінійного рівняння регресії: $b_0 = 1{,}5336, b_1 = 0{,}2407, b_2 = -0{,}0645$.

Моделі нелінійної регресії для оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate були побудовані за загальною формулою, з урахуванням одновимірної нормалізації методом Бокса-Кокса, що має вигляд:

$$\hat{Y} = \left(\left[b_0 + b_1 \cdot \frac{X_1^{\lambda_{X_1}} - 1}{\lambda_{X_1}} + b_2 \cdot \frac{X_2^{\lambda_{X_2}} - 1}{\lambda_{X_2}} \right] \cdot \lambda_Y \right)^{\frac{1}{\lambda_Y}}.$$

Модель виконується в межах метрик RFC[1,6; 13,6222], CBO[2,5714; 9,0952] та WMC[2,1905; 16,2500]. Також, для всіх моделей розрахований квадрат відстані Махаланобіса не має перевищувати критичне значення $\chi^2_{0{,}005,10} = 25{,}1882$.

Наступним кроком виконуємо розрахунок показників якості для побудованих нелінійних регресійних моделей. Аналіз їх точності та перевірка адекватності здійснюватиметься за допомогою критеріїв якості MMRE та PRED.

Результати оцінювання моделі RFC(CBO, WMC):

– Навчальна вибірка: $MMRE = 0{,}3215$, $PRED(0.25) = 0{,}5111$.

– Тестова вибірка: $MMRE = 0{,}2434$, $PRED(0.25) = 0{,}6800$.

Результати оцінювання моделі CBO(WMC, RFC):

– Навчальна вибірка: $MMRE = 0{,}1929$, $PRED(0.25) = 0{,}8000$.

– Тестова вибірка: $MMRE = 0{,}2504$, $PRED(0.25) = 0{,}6800$.

Результати оцінювання моделі WMC(RFC, CBO):

– Навчальна вибірка: $MMRE = 0{,}3586$, $PRED(0.25) = 0{,}5111$.

– Тестова вибірка: $MMRE = 0{,}2350$, $PRED(0.25) = 0{,}7200$.

Відповідно до отриманих результатів було визначено, що усі три моделі демонструють прийнятну якість за критеріями MMRE та PRED(0.25) для навчальної та тестової вибірок.

Модель RFC(CBO, WMC) показує помірну якість на навчальній вибірці та значні покращення на тестовій, що свідчить про її здатність добре узагальнювати дані.

Модель CBO(WMC, RFC) характеризується найкращими значеннями MMRE та PRED(0.25) серед трьох моделей, що робить її найбільш точною та стабільною.

Модель WMC(RFC, CBO), подібно до першої моделі, демонструє середній рівень точності на навчальній вибірці та кращі показники на тестовій, що також підтверджує її практичну придатність.

Висновки. В результаті роботи було побудовано моделі для оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate. Отримані значення критеріїв MMRE та PRED(0.25) продемонстрували здатність усіх трьох моделей забезпечувати достовірне оцінювання, що дозволяє ефективно застосовувати їх для оцінювання якості застосунків з відкритим кодом на Java, що створюються з використанням фреймворку Hibernate.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Sellers, S. (2025). *State of Java Survey & Report.* https://securityboulevard.com/2025/01/state-of-java-survey-confirms-majority-of-enterprise-apps-are-built-on-java/

[2] Bagnulo, M., Claise, B., Eardley, P., Morton, A., Akhter, A. (November, 2021). *RFC 8911*. https://www.rfc-editor.org/rfc/rfc8911.html

[3] Shyam, R., Chris, F. (1993). *Metrics suite for object oriented design.* https://www.eso.org/~tcsmgr/oowg-forum/TechMeetings/Articles/OOMetrics.pdf

[4] Aanchal, S., Kumar. (2013). Metrics for Software Components in Object Oriented Environments: A Survey. *International Journal of Scientific Research in Computer Science and Engineering*. https://isroset.org/pub_paper/IJSRCSE/ISROSET-IJSRCSE-00043.pdf

[5] Prykhodko, S. (2024). “Evaluating Quality of Software Systems by the Confidence and Prediction Intervals of Regressions for RFC, CBO and WMC Metrics. *WSEAS Transactions on Systems*, vol. 23, pp. 322–330. https://doi.org/10.37394/23202.2024.23.36

Pukhalevych A.V, Kovalenko V.V.

A nonlinear regression models for assessing the quality of open source Java applications created using the Hibernate framework

Abstract*. In this paper, using Box-Cox transformations, nonlinear regression models RFC(CBO, WMC), CBO(WMC, RFC) and WMC(RFC, CBO) were constructed. Empirical data from 70 projects available on GitHub were used in the construction.*

Keywords: *development quality assessment; nonlinear regression model; hibernate framework.*